DOCUMENT RESUME

ED 359 244 TM 020 013

AUTHOR Rafferty, Eileen A.; Treff, August V.
TITLE Issues in Constructing School Profiles.
PUB DATE Apr 93
NOTE 5p.; Paper presented at the Annual Meeting of the American Educational Research Association (Atlanta, GA, April 12-16, 1993).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; Administrator Role; *Comparative Analysis; Cutting Scores; Educational Improvement; Educationally Disadvantaged; Elementary Secondary Education; *Error of Measurement; Evaluation Problems; Measurement Techniques; *Profiles; Research Reports; Resource Allocation; Sampling; School Effectiveness; *School Statistics; Student Mobility; *Test Results; Transient Children
IDENTIFIERS Diversity (Groups); *Diversity (Student); Missing Data

ABSTRACT

Some of the issues faced by institutions as they attempt to design a system of constructs that reflects the diversity of their schools are addressed. The choice of statistic, percentage passing versus mean, used in reporting test results can impact the allocation of services to subpopulations in the school. Percent passing statistics tend to focus resources on students whose scores fall near the cutoff for passing, while annual comparisons of mean scores tend to be affected most by students scoring at the extremes of the distribution. Sample bias is another issue that cannot be ignored. Test scores can be affected by large numbers of geographical transients. Errors of omission can have a snowball effect on statistics such as attendance, and distortion attributable to omission can affect test results as well. Administrators must look beyond common sense indicators of school success to construct statistical profiles that reflect disparate populations fairly. Indicators must be designed so that community pressure to show improvement does not, in fact, reward or ignore deleterious practices. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
EILEEN A. RAFFERTY
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED359244

Issues in Constructing School Profiles

Eileen A. Rafferty & August V. Treff [1]

AERA Annual Meeting Presentation, April, 1993, Atlanta, GA

The necessity of meeting school-level accountability standards has prompted increasing numbers of state and local educational agencies to catalog statistical data in the form of school profiles. The demand for comprehensive documentation of educational strengths and weaknesses on a school-by-school basis has been fueled in part by the emergence of magnet school programs and by other occasions in which students and parents are allowed their choice among competing public institutions. What statistics should be included in such profiles? What factors may complicate or even distort the profiles? What unintended consequences may arise as a result of increased pressure to show yearly school improvement? This paper addresses some of the issues faced by institutions as they attempt to design a system of constructs that reflect the diversity of their schools. Although the statistical issues discussed below can apply to a number of school indicators, many of the examples used in this paper relate to reporting high-stakes test results.

1. ***Statistic Type*** - The choice of statistic, percentage passing vs. mean, used in reporting test results can impact the allocation of services delivered to subpopulations within the school. Once a test has been established within the district, a school's yearly test performance tends to be evaluated in terms of annual increases or decreases in score. Different score types can have dramatically different effects on how such increases are achieved.

Percent Passing statistics tend to focus resources on students whose score falls near the cutoff for passing. (Although more complicated, much of the reasoning in this section also applies to *median* scores which can be thought of as a metric delineating a "cutoff" score falling at the midpoint of the distribution.) Even significant improvements in the delivery of services to low-scoring students may not increase their scores enough to exceed a cutoff score set a dozen or more NCE points above their baseline performance. If not, the services allocated to these children will not impact annual comparisons of the percent passing score. Nor will services delivered to high achieving students. Raising the test scores of students who were expected to pass the test without special services does not impact the percentage of students passing a test.

In one inner-city district, regression equations were shown to successfully target students predicted to place at or near a passing score equivalent to the 58th national percentile. The test scores of this subpopulation, which included fewer than 20 percent of the students being tested, accounted for nearly all students scoring within four NCEs of the cutoff score. Further analysis revealed that no student receiving services for Chapter 1, English as a Second Language, and/or Gifted and Talented services scored within the bandwidth surrounding the cutoff score. Changes affecting these programs or the delivery of services to these children had no impact on increases or decreases in annual score comparisons.

Mean Scores - By contrast, annual comparisons of *mean scores*, although they reflect the performance of students at all levels of achievement, tend to be affected most by students scoring at the extremes of the distribution. The same Chapter I and Gifted students who may have little impact on increases or decreases in percent passing are the same students whose accomplishments or lack thereof can unduly influence increases or decreases in means. A

TM020013

change in policy or even chance variation which affects traditionally low- or high-scoring populations can inordinately influence mean score increases or decreases.

In addition to the dangers of skewed representation, means can promote the perception of a homogenized, stereotypical student population. Unlike the percent passing statistic which highlights the distinction between passing and non-passing students, mean statistics offer an undifferentiated composite which can disguise important information. In recent years, an average or mean score has come to be interpreted by some as a baseline or minimal standard. Such impressions have contributed to the "Lake Wobegone Effect" (Cannell, 1988; Linn, Graue, & Sanders, 1990) in which "all students score above average." Students and their families concerned with choosing a school to meet the unique needs of a particular student may not be best served by mean score reporting. Policy-makers and those concerned with evaluating the achievement of a heterogeneous school community might also require more than means.

2. ***Sample Bias*** - Favorable test scores may sometimes conceal contributing, but negative, factors. A high drop out rate or a tendency to retain slower students may sometimes inappropriately contribute to higher test scores. All else being equal, a school which successfully lowers its dropout rate risks lowering its test scores as numbers of low-scoring students are encouraged to complete their education. Similarly, a school which elects to promote, rather than retain, low-scoring students may also risk lowering its test scores (e.g., Slaven & Madden, 1991; Ligon, 1991; McGill-Franzen & Allington, 1993).

The practice of retaining low-scoring students effectively biases the sample on which achievement test scores are calculated. As a cohort progresses through school, the winnowing process, repeated each year, exacerbates this bias. Figure 1 shows the percentage of overage students at each grade level participating in a spring 1992 census-testing administration for the Baltimore City Public Schools. Retention, operationally defined in terms of student birth year relative to current grade level[2] (e.g., Smith & Shepard, 1987), shows a linear increase with the grade level assessed.

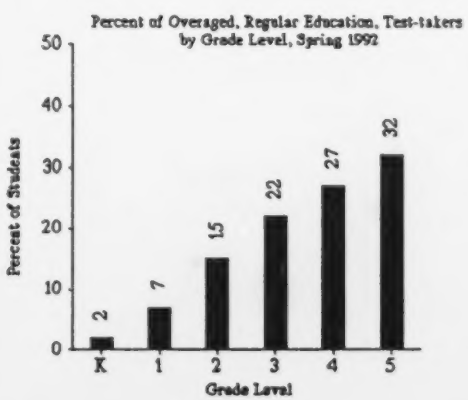

FIGURE 1. *Percentage of overage, regular education, test-takers grades K through five in Baltimore City. Public Schools as of spring 1992.*

The increasing disparity in the composition of the test-taking population across grade levels can lead to statistical anomalies which promote misinterpretation of test results. One such phenomena is the Simpson Paradox (Jaeger, 1992; Linn, 1993) in which

overall results decline while each of the major subgroups composing the overall figures show increases for the same period. Such results are dependent on unadjusted increases in the percentage of relatively low-scoring students taking the test. As the percentage of low-scoring, retained students increases with grade level, score comparisons across grade level may also be affected. As Figure 2 shows, both retained (overage) and not retained (not overage) populations showed score trends which were not reflected in the district scores of all students. Whereas district results remained constant across grades three to five, the scores of both retained and not retained students showed improvements between fourth and fifth grades, and the scores of overaged students in third grade declined relative to the scores of their counterparts in grade four.

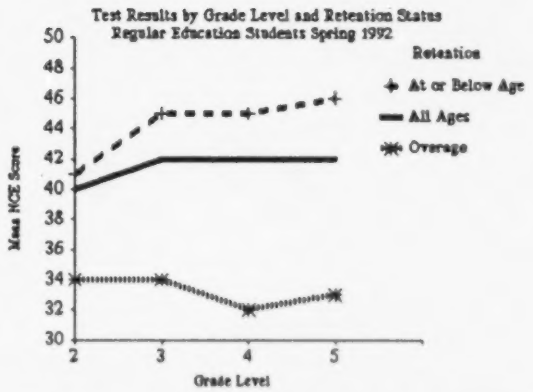

FIGURE 2. *Mean NCE scores of overage (retained), non-overaged (not retained) test-takers, and all test-takers grades two through five Baltimore City Public Schools as of spring 1992.*

Concurrent publication of the percent of students in each grade level who have been retained at least one year along with disaggregations of scores by retained and not retained students may help control for such distortions. Others have recommended the development of age appropriate rather than grade appropriate norms.

3. ***Geographical Transients*** - Schools in which a large number of students are transients who may or may not be proficient in English, can show understandably low test scores. Because of these issues, poor test performance may be routinely ignored by schoolbased staff and by district administrators who assume that the scores cannot be attributed to factors under the control of school personnel. In such cases, it may be advisable to disaggregate test results into transient and non-transient groups. In one urban elementary school, Metropolitan Achievement Test results of students who had transferred into the school during the year were compared with the scores of students who had attended the school for two or more years. The sixty percent of the population who were transient scored significantly higher than did students who had not transferred. It was hypothesized that lowered expectations had generalized to the entire school population.

4. ***Errors of Omission*** can have a snowball effect on statistics such as attendance. A optical scanning sheet had been used for several years by one district to collect attendance figures. School staff bubbled in grids for students who had attended each day. A survey of 178 schools over a six month period showed that a total of 3,935 students had monthly records which were

blank, indicating that the students had not attended classes for the entire month. On examination, 45 percent of these chronically absent students were located in just five percent the district's schools. These were schools who failed to keep current records.

Distortion attributable to omission can affect test results as well. Even when administering a high-stakes test, it was found that approximately two percent of one district's teachers had instructed their students to attempt only selected test items on a multiple-choice test. In each case, relatively large numbers of contiguous items within a subtest were left uniformly blank by every student in the affected classroom. When questioned, some teachers indicated that the unattempted items represented information which the teachers had not covered. Others expressed concern that taking the test would impair student self-esteem.

As administrators, we must look beyond "common sense" indicators of school success to construct statistical profiles which fairly reflect disparate populations. Furthermore, indicators should be designed to ensure that community pressure to show improvement does not, in fact, reward or ignore deleterious practices.

[1] Baltimore City Public Schools, 200 E. North Ave. Room 203, Baltimore, MD 21202

[2] Birth year algorithm used to estimate retentions: (School year as of September - grade - year of birth - 5).

References

Cannell, J. J. (1988). Nationally normed elementary achievement testing in America's public schools: How all 50 states are above the national average. *Educational Measurement: Issues and Practice, 7(2)*, 5-9.

Jaeger, R. M. (1992). "World Class" standards, choice, and privatization: Weak measurement serving presumptive policy. Address presented at the annual meeting of the American Educational Research Association, San Francisco.

Ligon, G. (1991). Models for Identifying Chapter 1 Schools for Improvement. *Educational Evaluation and Policy Analysis, 13*, 389-393.

Linn, R. L. (1993). Educational Assessment: Expanded Expectations and Challenges. *Educational Evaluation and Policy Analysis, 15*, 1-16.

Linn, R. L., Graue, M. E., & Sanders, N. M. (1990). Comparing state and district test results to national norms: The validity of claims that "everyone is above average." *Educational Measurement: Issues and Practice, 9(3)*, 5-14.

McGill-Franzen, A., & Allington, R. L. (1993). Flunk'em or Get Them Classified: The Contamination of Primary Grade Accountability Data. *Educational Researcher, 22(1)*, 19-22.

Slavin, R. E. & Madden, N. A. (1991). Modifying Chapter 1 Program Improvement Guidelines to Reward Appropriate Practices. *Educational Evaluation and Policy Analysis, 13*, 369-379.

Smith, M. L. & Shepard, L. A. (1987). What doesn't work: Explaining policies of retention in the early grades. *Phi Delta Kappan, 69*, 129-134.